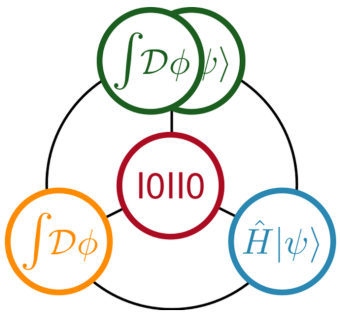
A05: Integrating HPC-Simulations with Data-Analysis for Structure Formation in Chemistry

Barbara Kirchner Petra Mutzel Estela Suarez

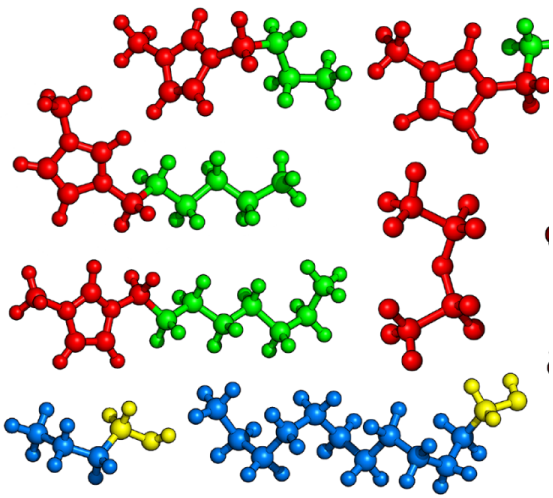
Mulliken Center for Theoretical Chemistry and Institute of Computer Science
University of Bonn

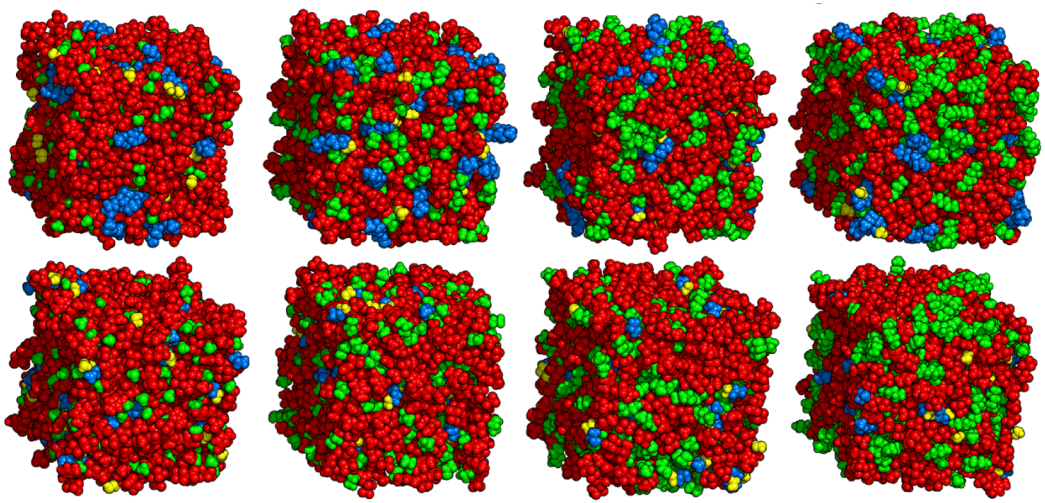


NuMeriQS Retreat 2024, Sep 30 – Oct 2, 2024

Overview

Aim of Project A05

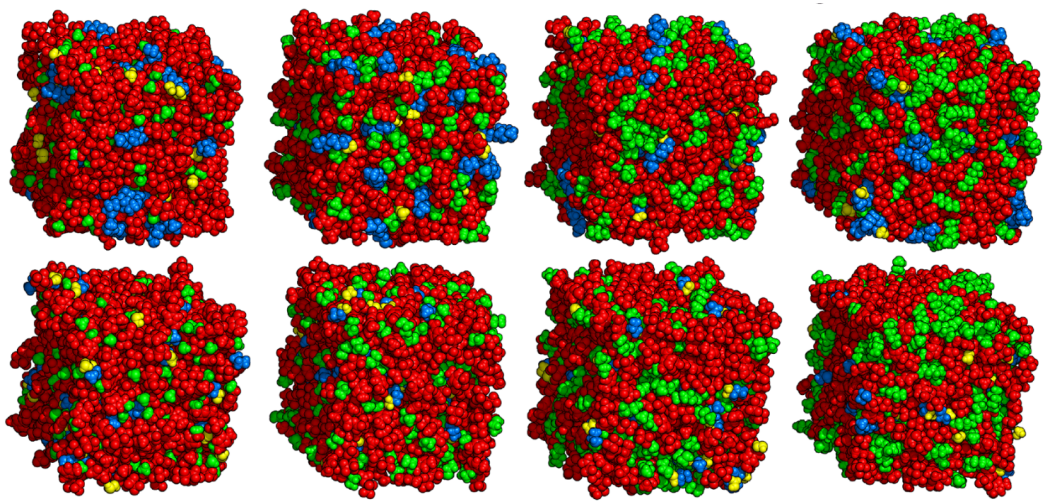




Deeper understanding of Molecular Systems

- evolution of conformations over time
- efficient use of computational resources enabling the investigation of more complex chemical problems
- integrated Ab Initio Molecular Dynamics (AIMD) simulations and data analysis

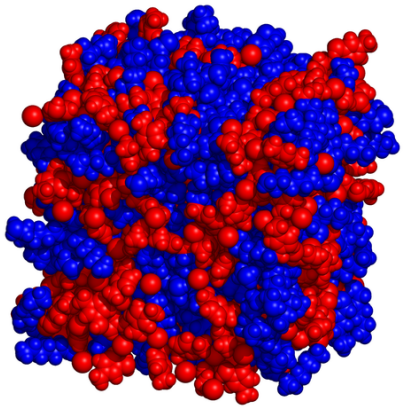
Summary of Project A05

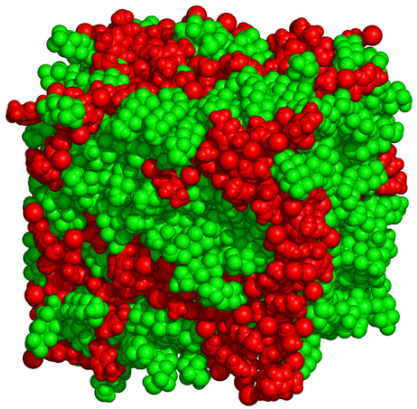


Approach

- run multiple AIMD simulations simultaneously and use novel (temporal) graph algorithms to exchange trajectories to lead the simulation
- develop analysis tools improving quality, efficiency, and scalability
- utilize the Modular Supercomputing Architecture (MSA) to match diverse workflow requirements

State of the Art

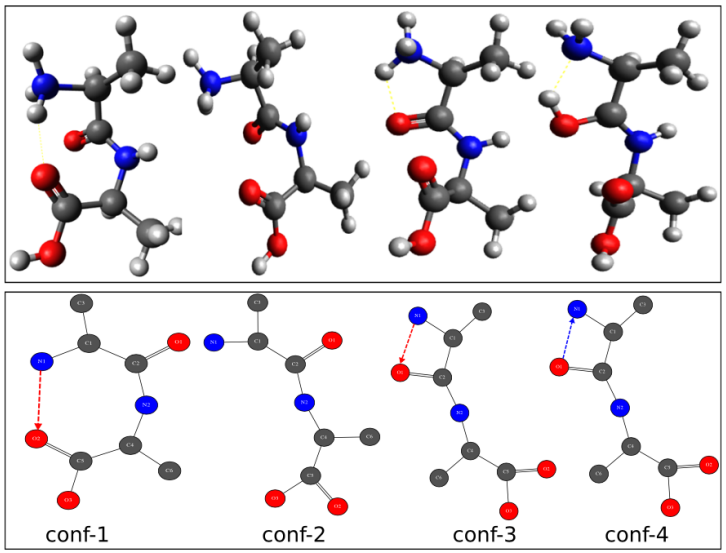


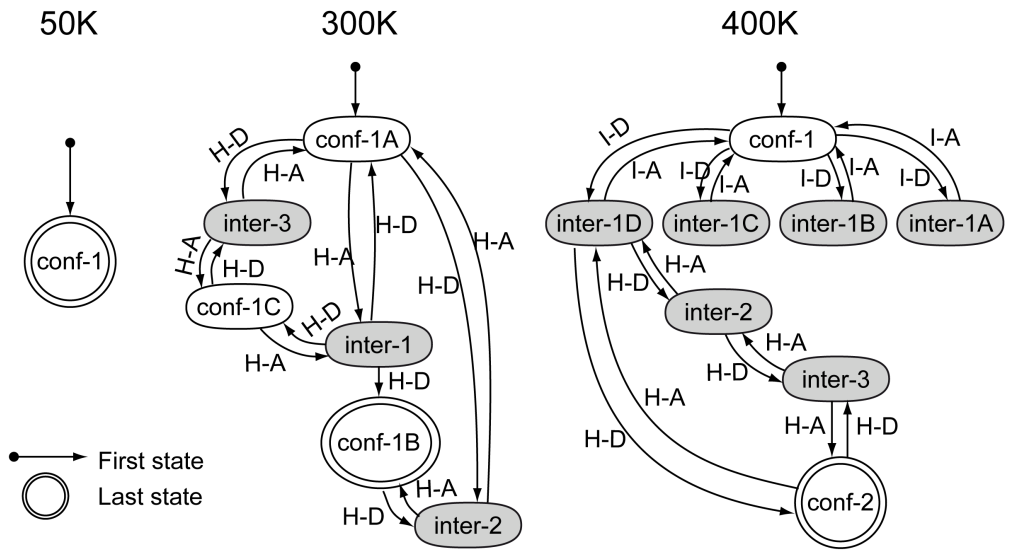


AIMD simulations

- enable study of structure formation effects in complex chemical systems, and vibrational spectroscopy such as Infrared, Raman, Vibrational Circular Dichroism (VCD)
- **Challenges:** optimal performance and scalability on large HPC systems

State of the Art



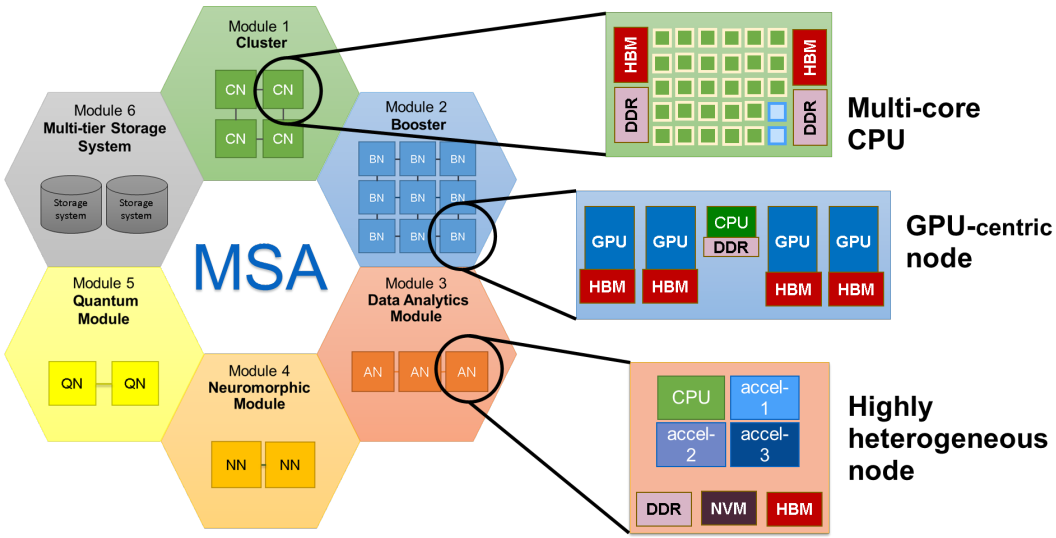


Graphs

- so far static structural formations and transition graphs used
- **Challenges:** lack time evolution information and struggle to recognize similar conformations

S. Bougueroua et al. 2018, Pietrucci et al. 2011

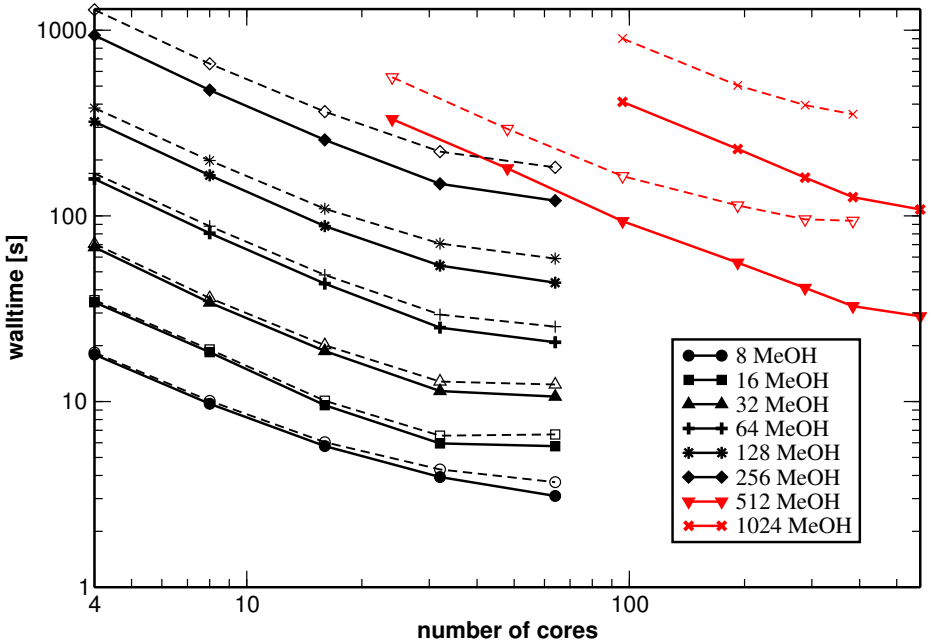
State of the Art



High Performance Computing

- AIMD simulations with standard parameters
- **Challenges:** efficiently use heterogeneous HPC systems by considering thread parallelism, memory management, and communication

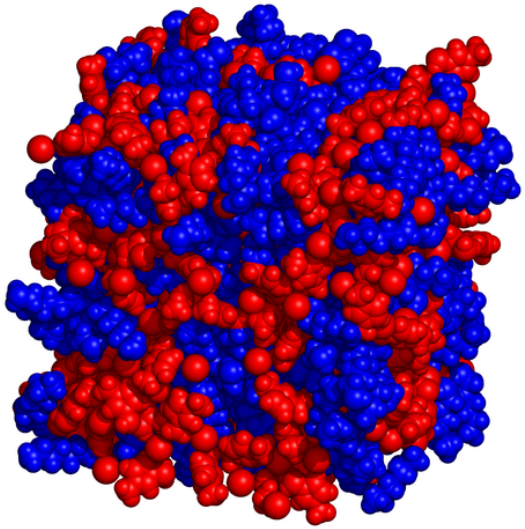
Preliminary Work: Scaling studies of AIMD with MSA

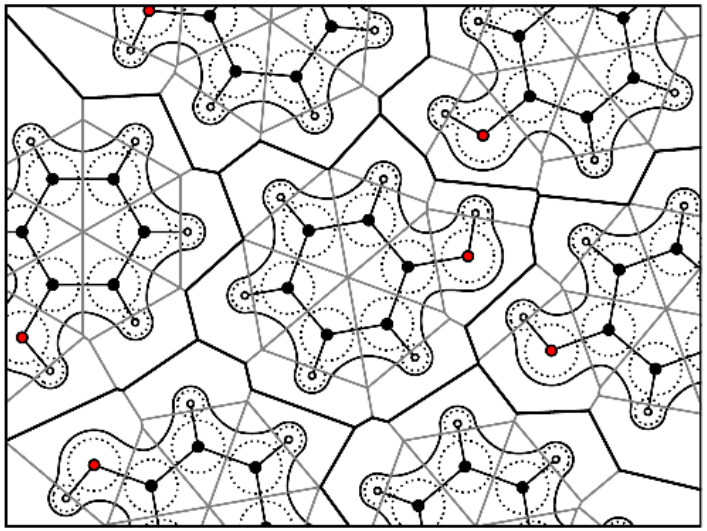


- MSA connects clusters with unique hardware configurations to meet user needs
- scaling studies on HPC systems at Jülich Supercomputing Centre

S. Taherivardanjani, et int., and **E. Suarez**, **B. Kirchner**, Benchmarking the Computational Costs and Quality of Vibrational Spectra from Ab Initio Simulations, Advanced Theory and Simulations, vol. 5, 2021

Preliminary Work: AIMD

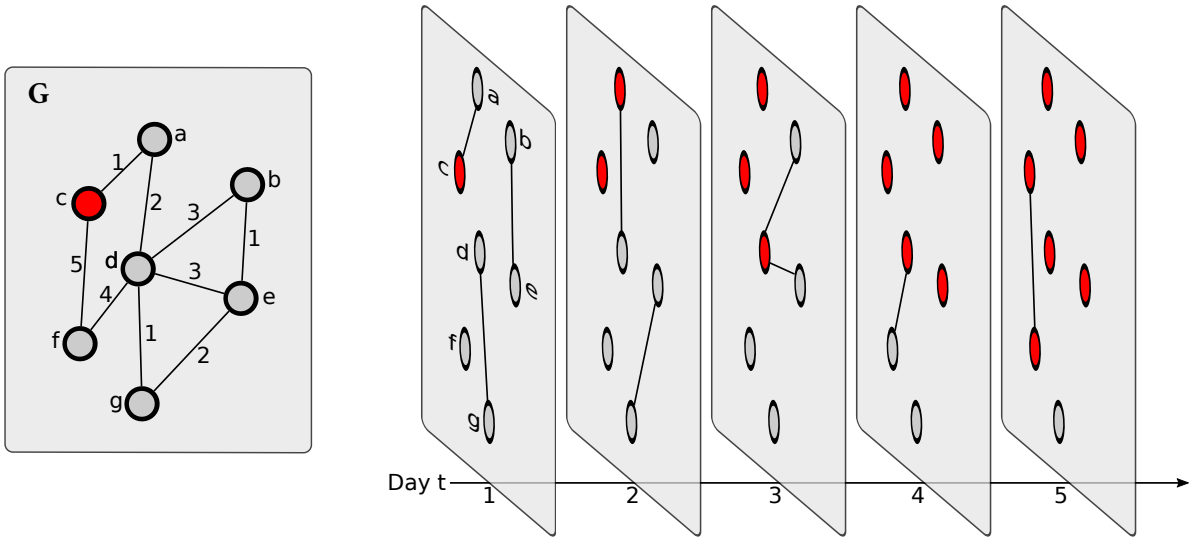




- trajectory analysis
- radical Voronoi tessellation for domain analysis and vibrational spectra calculation
- group atoms into subsets based on their local neighborhood

M. Brehm et int., and **B. Kirchner**: Domain Analysis in Nanostructured Liquids: A Post-Molecular Dynamics Study at the Example of Ionic Liquids, ChemPhysChem, vol. 16 (15), 2015

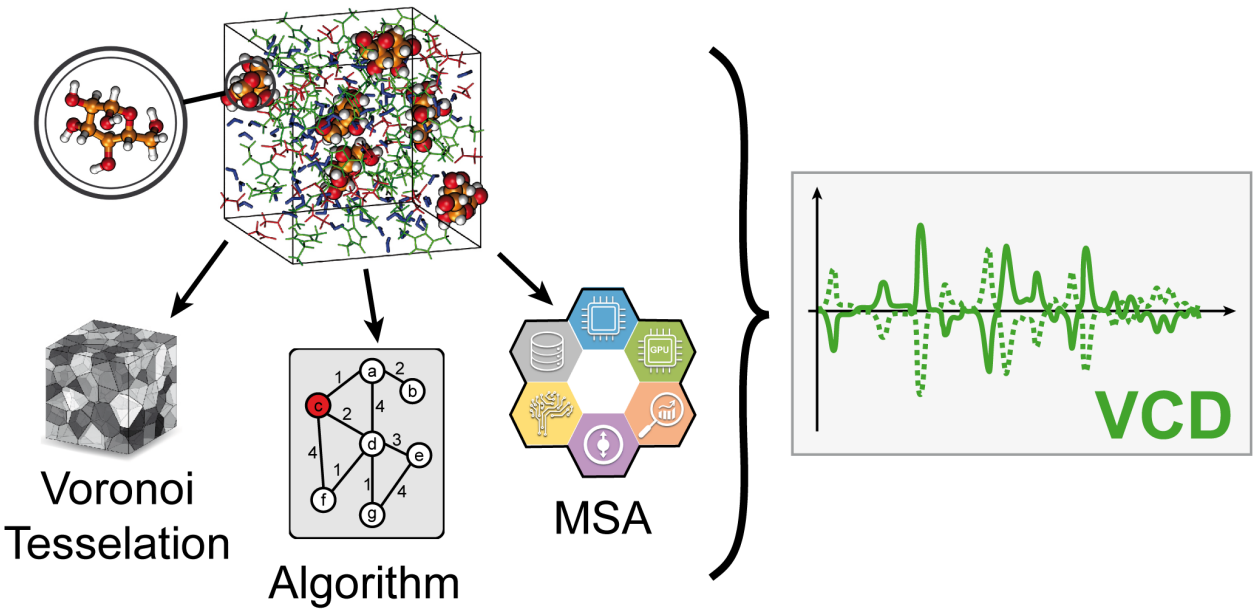
Preliminary Work: Temporal Graphs



- represent dynamic processes (since edges have a time stamp)
- new similarity measures and ML approach leading to improved classification rates
- broad experience in graph analysis and optimisation with applications in cheminformatics and theoretical physics

L. Oettershagen, et int., and **P. Mutzel**: Classifying Dissemination Processes in Temporal Graphs, Big Data, vol. 8 (5), 2020

Goals and Approach

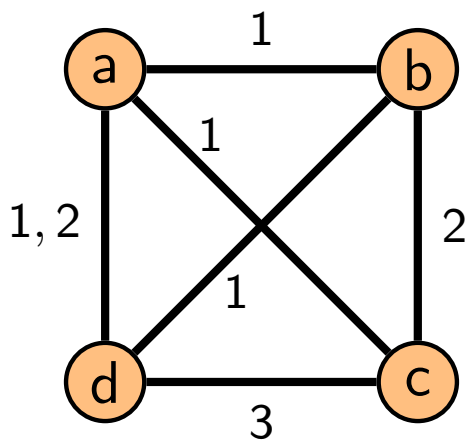


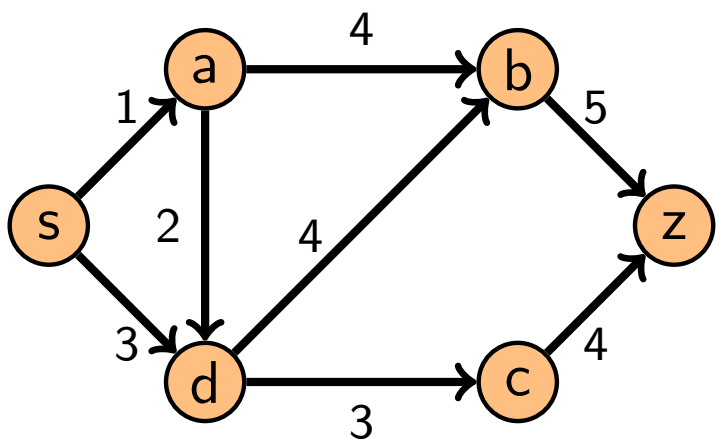
- **AIMD:** improve capabilities and performance of AIMD simulation and analysis of molecular vibrational spectroscopy
- **Graphs:** develop novel analysis methods and tools for dynamic processes based on temporal graphs
- **HPC:** improve MSA via co-design and adapt algorithms and codes to heterogeneous and modular supercomputing platforms

Temporal Graph Analysis

Definition (Temporal Graph)

- temporal graph $G_T = (V, E_T)$ has temporal edges $e_T = (u, v, t_e, \lambda_e) \in E_T$, where $u, v \in V$, $t_e \in \mathbb{N}$ and $\lambda_e \in \mathbb{N}$.
- Each edge e_T is present at time t_e.
- Traversal of edge e_T needs time λ_e.

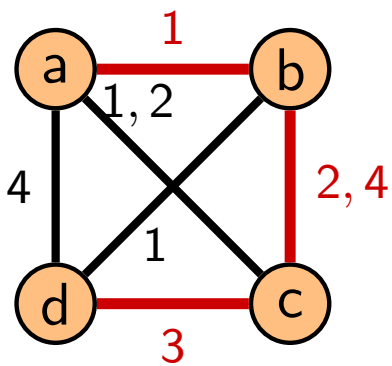


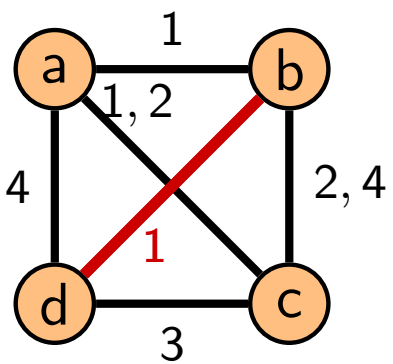


Two temporal graphs $G_T = (V, E_T)$ with $\lambda_e = 1$ for all $e \in E$

Temporal Paths and Walks

- Temporal paths in a temporal graph $G_T = (V, E_T)$ with temporal edges $e_T = (u, v, t_e, \lambda_e) \in E_T$, respect the time
- Find: minimum duration paths, earliest arrival paths, shortest paths
- Problem: subpaths of optimal paths not optimal

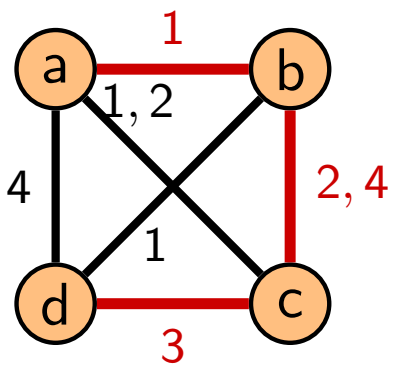


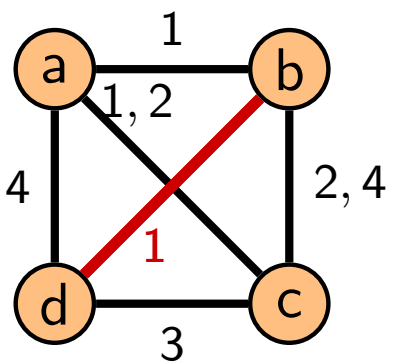


Earliest arrival paths from

Temporal Paths and Walks

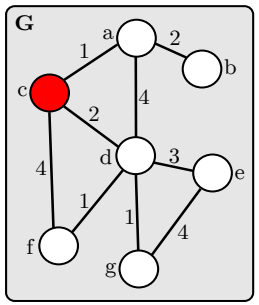
- Temporal paths in a temporal graph $G_T = (V, E_T)$ with temporal edges $e_T = (u, v, t_e, \lambda_e) \in E_T$, respect the time
- Find: minimum duration paths, earliest arrival paths, shortest paths
- Problem: subpaths of optimal paths not optimal

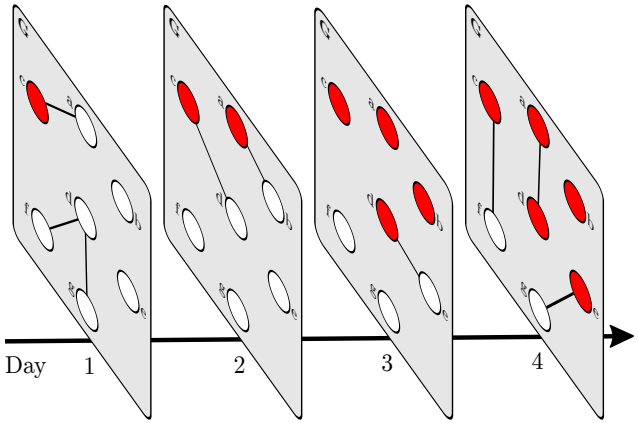




Earliest arrival paths from $a \to d$ and $b \to d$ ($\lambda_e = 1$ for all $e \in E$)

Applications



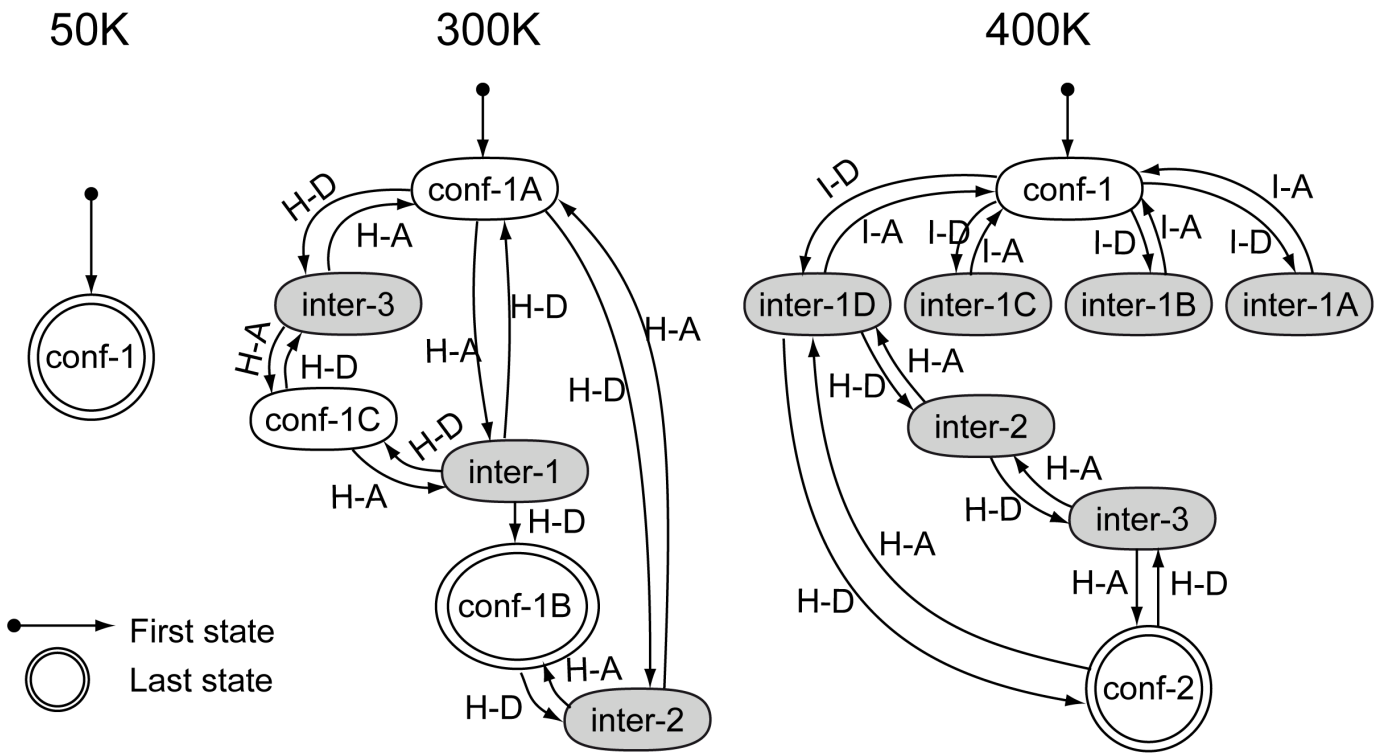


Temporal networks can be used to transfer information about dynamic processes to the network itself

Dissemination processes

- dissemination of fake news in social networks
- dissemination of infectious diseases
- modelling of dynamic processes in biological networks
- brain dynamics (e.g. from fMRI data)
- dynamics in chemical structure formation

Back to our Structure Formation Problem



Temporal Graph Analysis allows us:

- take the time evolution information into account
- from temporal graph similarity $\rightarrow$ guide the simulations during the runs

Current Status

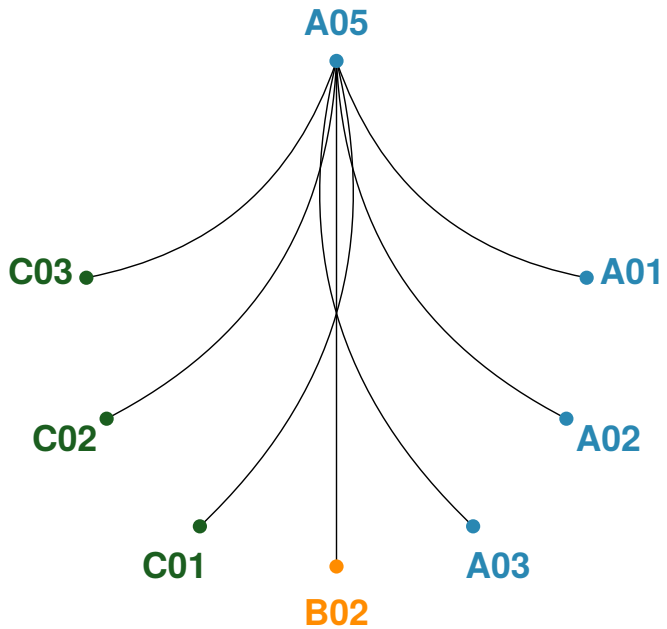
Staff situation

- Katrin Drysch (started on October 1st)
- cooperation with Ingo Scholtes (Univ. Würzburg)
- cooperation planned with Stefan Kesselheim (FZJ)

Ongoing Work

- process flow by Estela and Barbara
- joined Master thesis supervised by Barbara and Petra
- study temporal graph isomorphism based on De Brujn graphs

Role of A05 within the CRC



- graph algorithms and graph learning with **A01**
- numerical methods for quantum chemistry with **A03**
- techniques of HPC and MSA optimisations with **A02** and **B02**
- generalizations of used Delaunay trinangulation with **C01**
- link to **Z02** for benchmarking and testing on HPC and MSA systems